Research Report

Cognitively Based Assessment of, for, and as Learning: A Framework for Assessing Reading Competency

Tenaha O'Reilly

Kathleen M. Sheehan

June 2009

ETS RR-09-26

Cognitively Based Assessment of, for and as Learning: A Framework for Assessing Reading Competency

Tenaha O'Reilly and Kathleen M. Sheehan
ETS, Princeton, New Jersey

June 2009

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

This paper presents the rationale and research base for a reading competency model designed to guide the development of cognitively based assessment of reading comprehension. The model was developed from a detailed review of the cognitive research on reading and learning and a review of state standards for language arts. A survey of the literature revealed three key areas of reading competency: prerequisite reading skill, model building skill, and applied comprehension skill. Prerequisite reading skill is the ability to read text accurately and fluently. Model building skill is the ability to construct meaning from either decoded text or spoken language. Applied comprehension skill is the ability to use and apply the information contained in text for some particular purpose. The framework is discussed in terms of 7 key principles that have implications for the design of a modern assessment of reading.

Key word: Reading comprehension, assessment, learning, reading strategies, 21st century skills

Acknowledgments

We are grateful to Katie Arnold, Barbara Elkins, and Heather Nadelman for leading the assessment development for this project. We would also like to thank Malcolm Bauer, Randy Bennett, Kelly Bruce, Paul Deane, Michael Ecker, Rene Lawless, Zydrune Mladineo, Sarah Ohls, Peggy Redman, John Sabatini, Jana Sukkarieh, Waverely VanWinkle, Mike Wagner, and the following teachers from the Portland, Maine, School District: Trudy Brown, Lisa Hatch, Gail Hood, Karen MacDonald, and Maureen Tevanian.

Table of Contents

Introduction

For the better part of the 20th century, the United States has been recognized for its social mission, innovation, and economic prosperity. However, with the increase in global communication and global marketing, the United States is now facing enhanced competition from countries such as India and China. America's continued economic prosperity and its position in the global economy are also threatened by three internal problems: a changing economy, divergent skill distributions, and demographic trends leading to a more diverse population (Kirsch, Braun, Yamamoto, & Sum, 2007).

America's changing economy has created a demand for a more highly skilled labor force, as the jobs that will be created in the future, more and more, will require a college degree. According to Kirsch et al. (2007), however, the United States is currently experiencing a general deficit in critically needed skill areas such as reading and math, and these achievement gaps have remained stable over time. Changing demographics are also a contributing factor because populations that traditionally had relatively low levels of education and literacy are increasing most rapidly. As Kirsch et al. argued, these factors, if left unchecked, threaten to dramatically change the U.S. position in the world economy. Consequently, Kirsch et al. referred to these trends as America's perfect storm.

What, then, can be done to help prevent America's perfect storm from happening? At least part of the answer depends upon policy makers' ability to promote educational reforms that will enable all students to acquire the reading and higher level thinking skills needed to succeed in the 21st century. Over the past 10 years, federal policy makers have advocated change though the vehicle of high stakes testing. Federal legislation such as the No Child Left Behind (NCLB) Act (2002) is intended to promote change by making schools accountable for the reading and math skills of all of their students. While the effects of high stakes testing are highly contested (McNeil, 2000; Nicholas & Berliner, 2007), it is likely that the NCLB legislation, probably in modified form, will continue to influence education in the near future.

Education reform has also gained traction at the state level as states have begun to wrestle with the effects of local economic changes such as the loss of manufacturing jobs and the outsourcing of technical work to China and India. State reform efforts have included a greater emphasis on science and technology education and the development of new curricular standards

that explicitly address the kinds of high level skills (e.g., critical thinking, problem-solving) valued by modern-day business leaders.

A comprehensive system designed to address the concerns noted previously is described in Bennett and Gitomer (in press). This new approach is called cognitively based assessment *of, for* and *as* learning (CBAL). The CBAL approach involves first developing a cognitive model of domain competency in the targeted content area and then using that model to design tasks and activities that synergistically address three crucial aspects of educational practice: accountability assessment, formative assessment, and professional support. Successful implementation requires five key innovations.

1. Individual tasks are designed to assess the integrated use of the knowledge and skills specified in the competency model and to target aspects of students' performances that cut across different curricular units.
2. Administration of the accountability assessment is distributed across multiple testing sessions spaced throughout the school year so that (a) the importance of any one assessment and occasion is diminished, (b) tasks can be more complex and integrative because more time is available for assessment in the aggregate,[1] and (c) test results can be provided to teachers while time is still available to take appropriate instructional action.
3. Tasks and activities are designed to be learning experiences in and of themselves so that a strategy of teaching to the test is then an appropriate instructional activity.
4. Because all of the assessments are based on a common competency model, results from the various accountability assessments can serve as prior information for subsequent formative and diagnostic assessments.
5. Assessments consist primarily of computer-delivered constructed-response tasks scored via a combination of technologies, including both computer-assisted human scoring and automated scoring.

This paper describes an application of the CBAL approach to the domain of reading comprehension. The purpose of the paper is to describe the rationale and research base underlying the CBAL reading framework. The paper first provides an overview of the CBAL

reading competency model, followed by a discussion of the research base and key cognitive principles that were used to create the CBAL reading framework and competency model. The paper then concludes with a brief summary.

The Cognitively Based Assessment *of, for,* and *as* Learning (CBAL) Reading Framework

The CBAL reading competency model was created by performing a detailed review of the cognitive research on reading and a survey of state reading standards. The starting point for the model was based on the framework contained in the Rand Reading Study Group (RRSG) report titled *Reading for Understanding Toward an R&D Program in Reading Comprehension* (RRSG, 2002). The RRSG identified three important aspects of reading comprehension: the reader, the text, and the reading activity. Characteristics of the reader include the cognitive processes, skills, strategies, and knowledge that enable the reader to extract meaning from the text. Characteristics of the text include the collection of linguistic and content features that contribute to the ease or difficulty of extracting meaning from text. Characteristics of the activity include the purpose, goals, tasks, and the expected outcome(s) of reading. The interactions between the reader, the text, and the activity occur in the context of the larger sociocultural environment.

Reading is a complex process that requires the coordinated use of a large number of skills. In a broad sense, successful reading requires three general classes of skills: prerequisite reading skill, model building skill, and applied comprehension skill (see skills branch in Figure 1 and in the appendix for a detailed description of the skills branch). Prerequisite reading skill is the ability to accurately read text at a rate that is grade-level appropriate. It includes the ability to recognize familiar words (e.g., house) and the ability to decode unfamiliar words. This latter ability is particularly important for struggling readers as many printed words are unfamiliar. It is also important for on grade–level readers since many technical content terms are unfamiliar (e.g., mitosis), not to mention place and person names (e.g., Somalia, Shrek) and product labels (e.g., Zantec). Prerequisite reading skill also includes the ability to read text fluently with appropriate expression and intonation (i.e., prosody). It is important to note that prerequisite reading skills are bound to printed text and directly represent reading ability in its literal sense. Reading ability in this context means the ability to decode and accurately recognize printed text, not the ability to comprehend text (see model building skill in the next paragraph). In short, prerequisite reading skill is the foundational set of skills that enables readers to advance to the comprehension stage

of model building. Without adequate prerequisite reading skill, comprehension of printed text will suffer.

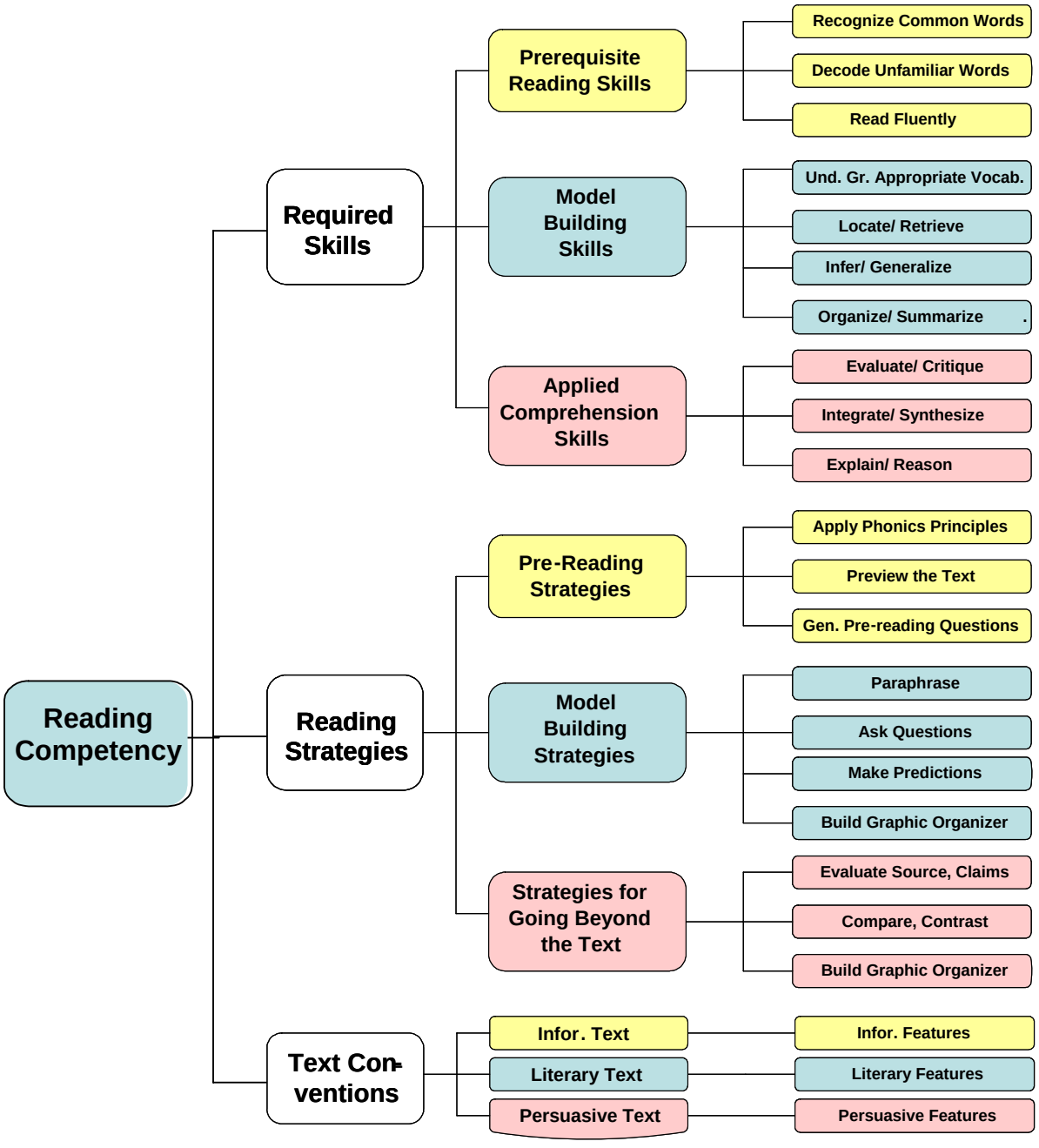

Figure 1. **The Cognitively Based Assessment *of, for* and *as* Learning (CBAL) reading competency model.**

Model building skill is the collection of abilities that allows one to construct meaning from either decoded text or spoken language. This skill set includes all of the skills needed to construct meaning from words (vocabulary), sentences, paragraphs, and the overall discourse structure of text. Model building involves the ability to locate and retrieve information (literal comprehension) as well as the ability to infer and generalize unstated relationships within text. Both the literal and inferential levels of text processing help the reader to construct a mental model of a text's meaning. A mental model is a structured representation of the literal and implied meaning of text. It includes the ability to chunk, organize, and summarize information. While model building skill can occur in the context of printed text, it is not necessarily tied to print; model building skill can operate in contexts other than print, such as spoken language.

Applied comprehension skill is the ability to use the information contained in text or spoken language for some particular purpose. Applied comprehension involves going beyond the literal and inferential interpretation of text or spoken language in order to use the information to achieve a particular goal (e.g., solve a problem, make a decision, create a presentation or Web site). Applied comprehension in the CBAL model is broken down into three types of reading: reading that requires integrating and synthesizing information from multiple sources; reading that involves reasoning, explaining, and generating explanations by integrating new information with relevant background knowledge; and reading that requires application of critical thinking skills to evaluate text contents (evaluate/critique).[2] Applied comprehension can occur in the context of printed text (reading) or in the context of spoken language; it is not bound to print. Moreover, applied comprehension may go beyond general definitions of reading comprehension by extending the definition to cover areas of critical thinking and critical reasoning.[3]

The competency model also includes the strategy branch to guide assessment design by ensuring tasks are included that call upon these strategies. The strategy branch is important because research has shown positive effects of reading strategy use on comprehension (McNamara, Ozuru, Best, & O'Reilly, 2007). Strategies are deliberate, conscious, effortful actions that successful readers implement to repair breaks in comprehension and to move understanding from a shallow level to a deeper level. In the CBAL framework, reading strategies are classified into one (or more) of the following categories: strategies that can help readers overcome deficiencies in prerequisite reading skill (e.g., previewing the text), strategies that can help readers build better (i.e., more coherent) mental models (e.g., representing text via tables or

other graphic organizers), and strategies that can help readers master the applied comprehension skill (e.g., adopt a critical stance, evaluate source/claims, looking for biases, etc.). Note that these classifications support the goal of integrating assessment and instruction by providing a link to the skill categories targeted in the assessment. In this way, the competency model provides a framework for defining assessment tasks that yield high quality evidence about targeted skills while simultaneously supporting both teaching and learning.

As is shown in Figure 1, the competency model also includes a text branch, which is the set of conventions and characteristics students must be conversant with if they are to be successful in making meaning from text. The text branch of the competency model represents the set of skills that students must have and which are accounted for in assessment design by including text with varying characteristics. Since many important text characteristics interact significantly with genre (see Sheehan, Kostin, & Futagi, 2008a, 2008b) distinct text difficulty models are employed for each of three targeted genres: literary text, informational text, and persuasive text. For more detailed information about these models, please see Sheehan et al.

Seven-Key Cognitive Principles and Their Implications for Creating a Cognitively Based Assessment of, for, and as Learning (CBAL) Accountability Assessment for Reading

The CBAL reading framework can be summarized in terms of seven key cognitive principles. This section provides a description, rationale, and research base for each principle. Implications for the design of CBAL reading accountability assessments that provide valid and reliable evidence about examinees' mastery status on targeted skills are also presented.

Provide a Realistic Purpose for Reading

Reading is an inherently purpose-driven activity (Alderson, 2000). Every day, people read for a variety of purposes, including to learn new information, to solve problems, to facilitate decision making, and to be entertained. Having a purpose for reading is important for several reasons, not the least of which is that the purpose helps to define the goals and motivation for reading. In essence, the reader's purpose provides a road map for determining which pieces of information are, and are not, important to focus on (Pressley, 2000). In other words, the purpose allows the reader to be selective in his or her reading. For example, if the goal of reading were to simply find out the year in which a specific historical event occurred, understanding (or even

reading) the sequence of actions that led up to that event would not be necessary as it would not directly help the reader achieve his or her specific goal.

Establishing a purpose for reading also helps to define the reader's standards of coherence (Linderholm, Virtue, Tzeng, & van den Broek, 2004; van den Broek, Lorch, Linderholm, & Gustafson, 2001; van den Broek, Risden, & Husebye-Hartman, 1995). The term *standards of coherence* refers to the level of comprehension that a reader strives to achieve. A reader's standards of coherence dictate how deeply a text is processed and the degree of uncertainty deemed acceptable. Readers who choose to have a high level of coherence must expend additional effort (e.g., elaborate, draw inferences) to ensure that the intricate details of a text are understood. Conversely, readers who establish a low standard of coherence are more likely to tolerate gaps in understanding.

Some evidence suggests that focusing a reader's attention on specific portions of a text can enable purpose-driven interest, which in turn promotes recall of the focused sections of the text (Schraw & Dennison, 1994); thus, providing a purpose for reading can increase interest and improve memory for text. This is an important result, as research has shown that intrinsic reader interest is positively related to reading comprehension (Cox & Guthrie, 2001; Wang & Guthrie, 2004).

It is no surprise, then, that research has shown that reading for different purposes can lead to differential processing of text (van den Broek et al., 2001) and differential use of reading strategies (Bråten & Samuelstuen, 2004). For instance, people who are asked to read for the purposes of studying for a test draw more coherence-based inferences than people who are asked to read for entertainment. Similarly, people are more likely to produce associations and evaluations when reading for entertainment than when reading to prepare for a test (van den Broek et al., 2001).

Despite the wealth of research on the impact of reader purpose on comprehension, many current assessments do not provide any purpose for reading other than to answer as many questions as possible in the allotted timeframe. As Rupp, Ferne, and Choi (2006) put it, "When people read in a nontesting context, they do not answer multiple choice questions in their heads" (p. 441). The take-home message then is that the contexts underlying most traditional reading assessments are somewhat arbitrary and artificial. Outside of the lab and testing situations, people read for specific and meaningful purposes. Perhaps the most important conclusion to

draw from the literature is that establishing a purpose for reading helps to improve reader interest, which in turn can improve comprehension.

The research summarized previously has implications for assessment design. In particular, it suggests that reading assessments should provide a goal-directed purpose for reading that is instantiated in the context of a meaningful scenario. The adjective *meaningful* implies that scenarios should be illustrative of the age-appropriate situations in which students would realistically be expected to read texts to achieve a particular purpose. Recent assessments that utilize scenarios in this way include the *Collegiate Learning Assessment* produced by the Council for Aid to Education and the *iSkills*™ test produced by ETS.

Scenario-based reading assessments contextualize the reading activity by first defining a meaningful purpose for reading and then presenting a collection of texts related to that purpose. Assessment scenarios should strive to achieve three aims: (a) provide a general reason for reading the particular collection of texts presented (e.g., to make a decision on whether to adopt a wind power as a source of energy in your community), (b) constrain the activity by providing specific subgoals (e.g., consider community-specific needs such as location and availability of wind), and (c) clarify expectations regarding the culminating task demands (e.g., create a Web site that informs the community about what wind power is, describes the pros and cons of building a wind farm, and provides possible solutions that address the drawbacks of using wind power as a source of energy in your community).

Satisfaction of these aims may also lead to additional benefits. For example, providing a reason for reading specific collection texts may also increase motivation and help students align their processing, resources, and strategies to match the task at hand. Providing subgoals allows the assessment designer to measure the ability to follow directions. Providing an expectation for what is to be produced should help increase student motivation because the reading activity is expected to produce some tangible result. Resulting scenario-based assessments not only provide a cohesive purpose for reading, but they also set the stage for the next two principles: integration and evaluation.

Measure the Ability to Integrate and Synthesize Information From Multiple Related Texts

Goldman (2004) noted that, because many traditional off-the-shelf reading assessments are designed to measure comprehension of individual passages in isolation, higher level skills such as integrating and synthesizing information from multiple texts are typically not addressed

(although the paired passages on the National Assessment of Educational Progress [NAEP] Assessment, the SAT® verbal section, and some state assessments are an exception to this).

In the modern world, it is rarely the case that all of the information needed to achieve a specific purpose is available within a single text. Scenario-based assessments, by contrast, incorporate a *research-based model* of literacy that assumes (a) readers have a specific question that they want to answer, (b) the answer to that question involves gathering information from multiple documents, and (c) comprehending that information involves integrating and synthesizing information in order to create what Perfetti, Rouet, and Britt (1999) called a *documents model.*

The documents model describes the reader's understanding of (a) the information presented in each individual text, (b) how the various texts are related, and (c) an interpretation or synthesis of the big picture implied by the collection of texts as a whole. Information that is consistent across documents is likely to be included in the documents model, while information that is discrepant is more likely to be excluded from the documents model. This type of document thinking is common in disciplines such as science and history (Goldman & Bisanz, 2002; Wineburg, 1998).

With the advent of technologies such as the Internet, the availability of information is now so great that it places new demands on readers. Documents on the Internet are written for different purposes and at different levels of quality. Readers in the 21st century must make sense of this information by combining it into meaningful and consistent documents models.

Research on multiple document learning is sparse, particularly compared to the wealth of research available on how readers comprehend isolated texts (Goldman, 2004). The research that has been conducted, however, has shown that reader purpose has an effect on how well readers integrate information from multiple documents. Wiley and Voss (1996) found that students were more likely to create a coherent documents model that integrated information from multiple texts if they were asked to explain rather than describe how an event occurred. Similarly, Britt and Sommer (2004) found that reading purpose and reading strategies affect how well readers integrate information across texts (documents model). Specifically, the experimenters found that requiring students to summarize a text before reading a subsequent text improved the quality of their documents models. The experimenters also found that the quality of the documents model improved when students were asked to answer macrolevel questions (e.g., *why* questions) as

opposed to micro-level questions (*detail* questions). Both findings support the idea that strategies directed at organizing and structuring texts also help readers to integrate information across multiple texts.

In addition, Britt and Sommer (2004) found that explicitly asking students to integrate texts into a single story was much more effective at enhancing comprehension than asking students to read for the purposes of answering comprehension questions. This result provides evidence that not all readers spontaneously integrate multiple documents if they are not instructed to do so. In sum, promoting integration, either through explicit instruction or by fostering organization and structure strategies, helps to improve readers' ability to integrate multiple texts.

We acknowledge that the activity of integrating multiple texts to create a documents model is very demanding for young readers (Goldman & Bisanz, 2002). For instance, research has shown that high school students do not spontaneously integrate information from multiple sources (Britt, Goldman, & Perfetti, 1999). The results of Britt and Sommer (2004) indicated this is also frequently true of college students. However, as noted previously, when given the proper context and instruction, even extremely young children are able to connect information across multiple documents in meaningful ways (Brown & Campione, 1996; Rowe, 1994), and some researchers have argued that adolescence is the ideal time to introduce multiple-document thinking (Goldman, 2004). We believe that introducing younger children to meaningful and age-appropriate tasks that require the integration of multiple documents will better prepare them for the tasks they will be required to perform later in life.

Although the previous discussion suggests that each assessment should include at least two related texts, it is also important to ensure that each text used in the assessment is long enough to have an identifiable structure. In an assessment context, a trade off exists between the number of texts (with an identifiable structure) that can be read and the amount of time available for testing. As the number of texts in an assessment increases, the length of any given text in the assessment must decrease due to time constraints. However, using short texts is problematic because texts of insufficient length have little or no identifiable text structure (i.e., a text's organizational pattern), a factor that is critical for reading because it aids in the recall and comprehension of text (Gernsbacher, 1990; Meyer & Wijekumar, 2007; Ozuru, Rowe, O'Reilly, & McNamara, 2008; Williams, 2007). While the number of texts included in a single assessment

will depend upon the purpose stated in the scenario and the text genre, the total number of texts will probably have to be limited to about four. Accommodating more than four texts in a single assessment would likely require using texts that are too short and have no identifiable structure.

An obvious, but important, additional design principle is that the texts included in the assessment must have some identifiable connection among them. The similarities among texts can include the discussion of similar topics, themes, principles, problems, descriptions, or critical analysis. Similarly, it is important that the texts included in each assessment be written by different authors and from different perspectives. Including texts that are too similar defeats the purpose of integration, which is to support cross-document thinking (and ultimately transfer of learning); if the documents are too similar, essentially they are a single document. Providing documents that are written from different perspectives may also help to promote a more liberal and sensitive attitude toward viewpoints held by different interest groups.

Finally, integration and synthesis questions can take on at least two forms, one in which the goal of the reading is to synthesize the information from all documents to form a single understanding of the common message across different texts (i.e., documents model). Alternatively, integration and synthesis questions can be somewhat less demanding by requiring the reader to select portions of the various texts in order to achieve a specific goal. In the first instance, the task essentially requires the reader to produce a common understanding for all of the documents. In the second case, the task requires the reader to select portions from the various texts to answer a particular question or achieve a particular purpose; this type of task may not necessarily require the construction of a common documents model. In sum, each assessment must contain multiple documents selected to support intertextual thinking.

Measure the Ability to Adopt a Critical Stance

With the advent of the Internet and other technologies, people have a seemingly limitless amount of information at their finger tips. Because no uniform standards exist for posting information on the Internet, however, the information found there can be unsubstantiated and inaccurate (Fritch & Cromwell, 2002; Metzger, Flanagin, Eyal, Lemus, & McCann, 2003). Now, more than ever, people need to approach all forms of print and nonprint-based sources of information with a *critical stance*. A critical stance can be defined as a disposition or tendency to "presuppose that the quality of [all] information is potentially suspect and requires scrutiny with respect to its truth, relevance, and other dimensions of quality" (Graesser et al., 2007, p. 90).

After surveying the research on critical evaluation skills, Metzger (2007) identified five criteria that can be used to evaluate information (primarily for expository and persuasive texts): accuracy, authority, objectivity, currency, and coverage. Accuracy concerns the truth of the information: Is the information valid and error free? Authority refers to the credentials of the author: What are the author's qualifications? Objectivity refers to the concept of bias and involves issues such as distinguishing fact from opinion, identifying conflicts of interest, and momentary gain: Is the author/source biased in his/her view? Currency refers to the time at which the information was published: Is the information out of date? Coverage refers to the amount of information contained in the document: Is there enough detail and depth? Is there missing information?

While Metzger's (2007) five-component model is useful in categorizing the important dimensions that readers should use in order to evaluate text, the model is incomplete because it does not consider the purpose of the reader. Accordingly, we suggest adding a sixth component to Metzger's model that includes the concept of relevance. Relevance refers to the degree to which the text information is pertinent for achieving the reader's specific goals: Is the information useful for answering the main question (Graesser, McNamara, & VanLehn, 2005; Graesser et al., 2007)? Such reflective, or metacognitive behavior is strategic and comprises an important part of learning (Azevedo & Cromley, 2004).

Evaluation, like integration, is a difficult, high level competency. Research has shown that students frequently approach reading in an uncritical fashion (Brem, Russell, & Weems, 2001; Greene, 1994; Vansledright & Kelly, 1998). For example, readers often fail to notice inconsistencies in text (Otero & Kintsch, 1992), fail to check for the accuracy of information (Flanagin & Metzger, 2000), and sometimes use a single criterion for making evaluations (Scholz-Crane, 1998). When evaluating online information, people will often use relatively superficial characteristics to make judgments of quality, such as the design and presentation of the Web site (Fogg et al., 2003).

Training explicitly designed to improve critical thinking skills has been shown to be effective at promoting a critical stance and learning (Britt & Aglinskas, 2002; Sanchez, Wiley, & Goldman, 2006). Now, more than ever, evaluation skills are becoming more important as the flow of available information will only increase over time.

The assessment of evaluation and critical thinking skills is particularly important for assessments that include persuasive text. Persuasive texts should include a range of content that varies in terms of the six dimensions mentioned previously—accuracy, authority, objectivity, currency, coverage, and relevance. This may require adapting texts to highlight important distinctions (e.g., adapting some texts to increase credibility and others to decrease credibility). Questions can then be designed to tap into a student's ability to evaluate texts based on the targeted dimensions. A similar approach can be employed to adapt expository texts when explicit signaling of potential problems is provided, and when it is developmentally and age appropriate. Questions should also be designed to ensure readers are effectively monitoring whether the information is relevant for their purpose.

Measure the Ability to Ask and Answer Deep Questions

Thus far, this section has focused on describing and providing the evidence base for three critical design features for the CBAL assessment: (a) providing a scenario that defines the purpose for reading, (b) providing multiple texts to facilitate integration and synthesis, and (c) providing texts that vary in terms of their quality to promote the use of evaluation skills. Collectively these three features are designed to promote deeper processing by encouraging readers to actively construct meaning from text (Singer, Graesser, & Trabasso, 1994). Unfortunately research suggests that readers often process text at a local and shallow level (McKoon & Ratcliff, 1992) and that such mindless processing can lead to an illusion of comprehension (Glenberg, Wilkinson, & Epstein, 1982; Langer, 1989; McNamara, Kintsch, Songer, & Kintsch, 1996).

At least two reasons explain why students process text at a superficial and minimal level: the context and the text itself. First, research has shown that in many subject areas (e.g., history and social studies), the curriculum emphasizes memorizing lists of facts rather than focusing on the deeper causal mechanisms driving the sequence of events (Smith & Niemi, 2001; Vansledright, 1995). Second, this fact-driven instructional approach is also reflected in the structure of many textbooks. In an analysis of social studies texts, Beck, McKeown, and Gromoll (1989) found that the content and organization of the texts did not encourage deep thinking. For example, Beck at al. found that texts often present too much information with too little detail (i.e., mile-wide, inch-deep), contain sets of loosely connected statements, and have poor integration with previous sections.

Taken together, both a fact-driven curriculum and poorly written textbooks place challenging demands on students who strive to achieve deeper levels of understanding. In order to develop a coherent model for relatively incoherent text, students have to explain how and why the ideas in the text are related. The act of explaining the meaning of text has been referred to as the *self-explanation effect* (Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; McNamara, 2004; McNamara, O'Reilly, Best, & Ozuru, 2006). Generating an explanation is an effortful activity in which readers construct meaning in order to draw inferences to fill in missing information, to make connections among ideas within the text, to integrate text information with existing knowledge, and to monitor and repair incorrect knowledge (Roy & Chi, 2005).

Research has shown that students who self-explain the meaning of text are more likely to solve problems successfully, draw important inferences, build more consistent mental models, and develop a deeper understanding of the material in the text (Chi et al., 1989; Chi et al., 1994). Other work has indicated that the number and quality of explanations are related to the reader's overall level of comprehension; students who produce more explanations, and students who produce higher quality explanations, learn more from text (Chi, 2000; McNamara et al., 2006). While some research suggests that high ability students produce higher quality and more sophisticated explanations than less skilled readers (Ozuru, Best, & McNamara, 2004), other work has indicated that specific training on how to self-explain primarily benefits lower but not higher ability students (McNamara, 2004; McNamara et al., 2006; O'Reilly, Best, & McNamara, 2004). The important conclusion to draw from this work is that while the process of generating explanations may be natural for higher ability students, lower ability students can learn how to generate better explanations through training.

In part, self-explanation is effective in promoting comprehension because it encourages readers to use their general knowledge to help fill in unstated relationships within the text (McNamara, 2004). In line with this reasoning, McNamara found that explicit training in self-explanation techniques helped reader's comprehension only for more difficult low-cohesion text; the training did not help comprehension for easier, high-cohesion text in which the relationships among the concepts were more explicit. An alternative but not contradictory view is that the process of generating explanations helps readers to identify and repair gaps in their understanding by making the process of comprehension more explicit and observable (see Chi,

2000, for a discussion of the incomplete text or repair view of self-explanation). In sum, the ability to generate explanations is a critical reading competency because it facilitates deeper and more complete understanding of text.

Although the research base suggests that the ability to explain text is an important reading competency, measuring a student's ability to do so poses significant challenges. First, generating self-explanations requires the reader to produce constructed responses. In some cases, the constructed responses can be lengthy, and as such, the production of extended responses shifts the construct more towards writing than reading. Because a separate CBAL writing strand already exists, requiring students to produce lengthy written explanations may not be appropriate. Second, scoring constructed-response explanations requires a considerable amount of human effort, including a significant amount of time devoted to training raters and scoring papers (Magliano et al, 2002; Magliano, Wiemer-Hastings, Millis, Muñoz, & McNamara, 2005; McNamara, Levinstein, & Boonthum, 2004; McNamara et al., 2006; Millis, et al. 2004). While automated approaches for scoring self-explanations have met with some success, the level of accuracy achieved may not be sufficient for use in high stakes settings (McNamara et al., 2004).

One possible solution to this problem is to create tasks and questions that load more heavily on carefully designed selected-response formats and, in some cases, short-answer constructed responses. Following the recommendations put forth in a recent *Institute of Education Sciences* (IES) report titled *Organizing Instruction and Study to Improve Student Learning* (Pashler et al., 2007), we advocate the design of assessment tasks that foster a student's ability to construct explanations by both asking and answering deep questions. According to the panel's recommendation, questions should focus on assessing causes, motivations, supporting evidence, planning, and logical justifications. Such evidence can be gathered by asking questions with the following stems (Pashler et al., 2007): Why? What caused X? How did X occur? What if, what if not? How does X compare to Y? What is the evidence for X? Why is X important?

Similarly, question generation could be evaluated using the framework described in Graesser and Person (1994), which provided a 16-level classification system for evaluating the quality of questions. In this classification system, lower level questions are those that assess simple concepts such as verification and providing examples, and higher level questions are those that assess enablement and interpretation. In sum, we recommend assessing the construct of explanation by requiring students to both ask and answer deep questions.

Measure Basic Model Building Skill

The skills discussed previously (i.e., integration, evaluation, and explanation) all belong to the CBAL category of applied comprehension skill. In terms of reading theory, applied comprehension skill is more closely associated with the situation model of representation because it requires the integration and application of the text information with existing knowledge (Kintsch, 1998). In this section, we discuss an additional collection of skills called *model building skill*. These skills are more in line with Kintsch's idea of a textbase representation that embodies the basic meaning, or gist, of a single text.

It is useful to distinguish four specific model building skills: the ability to understand grade appropriate vocabulary; the ability to infer and generalize information; the ability to identify, locate, and retrieve important details; and the ability to extract the discourse structure (chunk, organize, summarize). These four skills are described as follows.

Understanding grade appropriate vocabulary. It should be no surprise that knowing the meaning of the words in text is an important part of reading comprehension. Research has shown that a strong relationship exists between vocabulary and comprehension (Cunningham & Stanovich, 1997; Daneman, 1988; Hirsch, 2003). Correlations between comprehension and vocabulary measures are typically within the .6 –.7 range (Anderson & Freebody, 1981). While it is possible to comprehend a text without knowing the meaning of a few words, the role of vocabulary in comprehension should not be underestimated. For example, Nagy and Scott (2000) estimated that adequate reading comprehension depends on a person already knowing between 90 and 95 percent of the words in a text. Moreover, differences in vocabulary size can magnify and cause problems over time. Students who become competent readers typically have much larger vocabularies than struggling readers in first grade, and this advantage tends to grow dramatically over time (Graves, Brunetti, & Slater, 1982; Hart & Risley, 1995). But what does it mean to know a word? It is important to note that knowing the meaning of a word takes place on many levels. Words have intricate meanings, including their connotative and denotative functions. As such, one must consider not only the number of words a person knows (breadth), but also the depth at which those words are known (Ouellet, 2006; Tannenbaum, Torgesen, & Wagner, 2006). In sum, vocabulary development is a critical part of learning to read well and appears to be a significant aspect of the gap between competent and struggling readers. While the

existence of this gap is discouraging, a number of studies have demonstrated that teaching words directly can improve comprehension (Beck & McKeown, 1991; Stahl & Fairbanks, 1986).

Drawing inferences and making generalizations. The ability to draw inferences from text is one of the hallmarks of reading comprehension (McNamara & O'Reilly, in press). "An inference is defined as any piece of information that is not explicitly stated in the text" (McKoon & Ratcliff, 1992, p. 440). While a number of different types of inferences exists, the class of inferences that we are referring to here is more in line with the type of inferences required to build a coherent textbase, rather than the types of elaborative inferences that are characteristic of the construct of explanation (see Singer & Remillard, 2004). The ability to draw inferences is necessary because authors often fail to make the relations between concepts in the text explicit (Beck et al., 1989; Chi et al., 1994; Wilson & Anderson, 1986). As with vocabulary, the ability to draw inferences should not be understated, as it is one of the characteristic features that distinguish skilled from less skilled comprehenders (McNamara & O'Reilly, in press).

For instance, skilled comprehenders make more appropriate inferences at both the local and global level. At the local level, less skilled comprehenders have difficultly resolving referential information such as anaphors (e.g., knowing that *she* refers to *Mary*; Oakhill & Yuill, 1986; Yuill & Oakhill, 1988a). Less skilled comprehenders have increased difficulty with anaphor resolution when the distance between the antecedent and the referent increases (e.g., when the inference requires bridging over multiple sentences; Yuill & Oakhill, 1988a). Difficulties in relating information also extend to more global relations (i.e., generalizations). For example, less skilled comprehenders have difficulty relating successive topics in the text (Lorch, Lorch, & Morgan, 1987) and integrating information to construct the overall theme or main ideas of a passage (Palincsar & Brown, 1984; Stevens, 1988; Williams, 1986). On the other hand, skilled comprehenders are more likely to generate inferences that repair conceptual gaps in the text, whereas less skilled comprehenders are more likely to ignore gaps or fail to make the inferences to repair coherence gaps (Garnham, Oakhill, & Johnson-Laird, 1982; Long, Oppy, & Seely, 1994; Magliano & Millis, 2003; Magliano, Wiemer-Hastings, Millis, Muñoz, & McNamara, 2002; Oakhill, 1984; Oakhill & Yuill, 1996; Oakhill, Yuill, & Donaldson, 1990; Yuill, Oakhill, & Parkin, 1989). Generating needed inferences is a complex process that can break down at a number of different levels (Cain, Oakhill, Barnes, & Bryant, 2001). While the process of generating inferences is complex, it appears that explicit training on how to generate

inferences is an effective method for promoting comprehension (McNamara, 2004; Stevens, 1988; Williams, 1986; Yuill & Oakhill, 1988b). In short, the ability to draw inferences is a critical component of comprehension skill.

Identify, locate, and retrieve important details (literal comprehension). While the ability to infer information from text is an important reading competency, it is essential not to forget a more basic comprehension competency: the ability to understand the literal meaning of text. The literal meaning concerns the ideas, concepts, principles, events, dates, names, places, and people that are explicitly mentioned in the text; no inferences are necessary to understand literal information. In most comprehension assessments, students are allowed to reread the text while they are answering the comprehension questions. Under these circumstances, the ability to understand literal information essentially translates into the ability to locate and retrieve formation.

According to Guthrie and Mosenthal (1987), searching for information in text may be viewed as a problem-solving activity that includes forming goals, selecting portions of the text, making a decision as to whether the selected information satisfies the goal, remembering the information, and repeating the process until the goal is met. Not surprisingly, research has shown that skilled comprehenders are more efficient and accurate than less skilled comprehenders at locating information in text (Cataldo & Oakhill, 2000). Cataldo and Oakhill found that good comprehenders often go directly to the target information while poor comprehenders are undirected in their search. The ability of good comprehenders to locate information was not due to their general spatial ability or their ability to remember the specific spatial locations of individual words. Rather, good comprehenders were more likely than less skilled comprehenders to remember the in-text sequence in which specific words appeared. In addition, the experimenters found that the performance of good comprehenders was more akin to the less skilled comprehenders when reading a scrambled text, suggesting that the good comprehenders took advantage of the coherent text structure to help them locate information. Together, these results suggest that the ability to locate information in text is dependent both on the structure of the text and the reader's representation of the material.

The idea of measuring vocabulary, inference, and literal information is not new to the assessment community. Most existing comprehension assessments measure most if not all of these constructs. Together, these constructs are meant to sample and approximate the reader's

textbase representation of the material (the reader's model building skill). However, approximating the textbase by only measuring its components is problematic. The textbase is an integrated representation of the text content that exists in the reader's mind (Kintsch, 1998), and the whole is more than the sum of its parts. The point to be made is that if the goal is to measure the reader's representation of the text, then a measure that more directly reflects the quality of a reader's mental representations is needed.

Extracting the discourse structure—The ability to chunk, summarize, and organize information. While measuring vocabulary, inference, and literal comprehension skills is important, it is also important to measure examinees' performance relative to the skill of extracting the discourse structure. Two lines of research hold promise for facilitating this goal—summary writing and the work on text structure and graphical representations.

A summary is a concise representation of the text content that includes an extraction of main ideas and important details and how they are related to each other. Research has shown that the act of producing summaries results in better comprehension (Moore, 1995), retention of text information (Rinehart, Stahl, & Erickson, 1986), and in some cases, better course performance (Radmacher & Latosi-Sawin, 1995). Summary writing has also been shown to result in increases in students' ability to monitor their comprehension (Thiede & Anderson, 2003).

Not surprisingly, research has revealed that the quality of student summaries varies as a function of age, experience, and ability (Garner, 1985; Head, Readence, & Buss, 1989). For example, summaries written by less skilled comprehenders often lack consistency (McCutchen & Perfetti, 1982) and fail to adhere to the organizational structure of the text (Meyer, Brandt, & Bluth, 1980). Recent advances in automated scoring have enabled researchers to build programs that teach students how to improve their summary writing. One such system, called Summary Street, has been shown to improve both the quality of students' summaries and their reading comprehension (Franzke, Kintsch, & Caccamise, 2005).

Including the construct of summary writing in the CBAL reading assessment is problematic, however, because it requires the reader to construct a potentially lengthy written response. In other words, writing summaries in a constructed-response format may be more in line with the construct of writing than reading. This idea is supported by the fact that the quality of student summaries have been found to vary as a function of writing ability (Head et al., 1989). One possible solution to this problem is to assess summarization skills via a selected response

format. This approach is currently being used in the Test of English as a Foreign Language™ (TOEFL®). In particular, the TOEFL summarization task asks students to select appropriate summary statements from an option list that includes both correct and incorrect summary statements. Two types of incorrect summary statements are typically provided: statements that focus on minor details that are correct yet not appropriate for inclusion in a high level summary, and statements that represent erroneous inferences.

Research has also shown that the quality of students' summaries is frequently affected by structural characteristics such as headers. Topics that are signaled via headers are more likely to be included in student's summaries, while topics that are not signaled are less likely to be included (Lorch, Pugzles-Lorch, Ritchey, McGovern, & Coleman, 2001). This leads us to our second possible approach for assessing the reader's model of a text: the research on text structure and graphic representations.

Text structure is the organizational pattern that logically connects the ideas and relationships within text (Williams, 2007). Both expository and narrative texts have identifiable structures. In narrative text, the structure is often referred to as story grammar and includes elements such as the setting, characters, plot, conflicts, and resolution (Mandler & Johnson, 1977). Typical expository structures include comparison, problem and solution, cause and effect, sequence, description, and listing (Meyer & Wijekumar, 2007). Research has shown that texts with identifiable structures are easier to comprehend (Cataldo & Oakhill, 2000; Lorch et al., 2001). Other work has revealed that skilled readers often use the structure of a text to aid comprehension and recall, but less skilled readers are less likely to do so (Meyer et al., 1980). Furthermore, interventions aimed at teaching students to identify text structure have been shown to be effective for improving comprehension (Meyer & Poon, 2001; Meyer & Wijekumar, 2007; Williams, 2007). In sum, text structure is an important aspect of comprehension and provides an additional way to capture what we mean by an integrated representation of the reader's textbase (i.e., discourse structure[4]). Following, we briefly describe approaches for graphically representing the discourse structure.

As mentioned previously, we have concerns with using extended constructed-response formats for measuring the discourse structure. One alternative is to use graphical representations that map out the structure. In the research and educational fields, the use of graphic representations to organize text has been given a variety of different names including graphic

organizers, concept maps, and knowledge maps (Griffin & Tulbert, 1995; Lambiotte, Skaggs, & Dansereau, 1993; Robinson & Skinner, 1996; Vitale & Romance, 2007). A graphic representation can be a figure, diagram, table, or chart that is used to depict the hierarchical, categorical, sequential, causal, and logical relationships among the ideas in text. Unlike a summary, graphic representations make minimal use of writing because the focus is more on representing the organization of the text's important points rather than on presenting a coherent description of the information. Graphic representations allow the reader to quickly see the important ideas and how they relate, making them an economical way to represent the discourse structure (Robinson & Skinner, 1996). Indeed, research has shown that graphic representations can improve recall, comprehension, and concept acquisition, and also help to improve the ability to efficiently locate information in text (Griffin & Tulbert, 1995; Lambiotte et al., 1993; Robinson & Skinner, 1996; Vitale & Romance, 2007). Examples of the different types of graphic representations may be viewed at http://www.cast.org/publications/ncac/ncac_go.html. In sum, we advocate measuring four components of model building: understanding key vocabulary, drawing necessary inferences, identifying important details, and extracting the discourse structure.

Measure the Adequacy of Fundamental Reading Skills

Thus far we have elaborated on the set of skills that is important for comprehension, but we have not yet discussed the collection of fundamental skills that enable students to read printed text. We refer to these fundamental reading skills as *prerequisite reading skill*. According to the simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), reading comprehension depends on both language processes (e.g., model building and applied comprehension skill) and word recognition and decoding processes. Furthermore, the model states that both skills are necessary for reading comprehension and neither skill by itself is sufficient. While the simple view of reading was developed over 20 years ago, the basic premises of the model still hold today (Cutting & Scarborough, 2006; Vellutino, Tunmer, & Jaccard, 2007). Research has shown that lower level reading skills such as decoding contribute unique variance in standardized reading comprehension test scores over word recognition and language measures (Cunningham, Stanovitch, & Wilson, 1990; Hoover & Gough, 1990). However, as word recognition becomes more fluent and automatized, listening comprehension becomes a stronger predictor of reading ability, though word recognition continues to contribute significant

variance even in skilled readers (Carver & David, 2001; Gough & Walsh, 1991; Haenggi & Perfetti, 1994; McCormick, 1994). Nonetheless, weaknesses in fundamental reading skills such as word recognition and decoding can impair a reader's ability to form an accurate model of a text (Perfetti, 1985).

Why are lower level reading abilities such as word recognition and decoding important for reading comprehension? According to Perfetti's verbal efficiency theory (Perfetti, 1985), reading, like any other skill, is subject to resource limitations. In order to perform any complex skill, certain basic processes must be automatized to free up resources for more demanding higher level processes. When basic skills are lacking, they draw valuable resources from the higher level processes that demand them.

In an extension of the verbal efficiency theory, Perfetti and Hart (2002) proposed the lexical quality hypothesis (LQH). According to this model, the quality of a word representation can vary on three levels: orthographic, phonetic, and semantic. The orthographic code specifies the word's spelling, the phonetic code determines the pronunciation, and the semantic code provides the meaning. The central claim of the LQH is that if one or more representations of the word code are deficient, then the quality of the overall word representation decreases yielding subsequent decreases in both the efficiency and accuracy of word retrieval operations.

Consequently, the LQH asserts that skilled and less skilled readers differ in two respects. First, skilled readers have more resources to repair or embellish impoverished representations. These resources may include more effective decoding, spelling, and grammatical skills. However, the key characteristic that distinguishes skilled from less skilled readers is the number of high quality word representations. Skilled readers are more efficient and accurate at retrieving words because the majority of their word representations are complete. This allows for fast and efficient retrieval of orthographic, phonemic, and semantic information. When word recognition and decoding become automatic, it frees up the resources needed for higher level processes such as inferencing (Cain, Oakhill, & Bryant, 2004; Daneman & Merikle, 1996; Perfetti, 1985). In other words, fast and efficient word identification skill is the foundation for text comprehension (Perfetti, 1985).

In addition to word recognition and decoding ability, the research base also points to fluency as a crucial skill. Recent work has demonstrated that the ability to fluently read connected text (as measured by rate and accuracy) is correlated with reading comprehension

ability as measured by standardized reading comprehension assessments (Cutting & Scarborough, 2006; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003), state accountability assessments (McGlinchey & Hixon, 2004; Wiley & Deno, 2005), and NAEP (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). In fact, Cutting and Scarborough found that a measure of reading fluency (rate) predicted 1% to 6% of additional variance of several standardized comprehension assessments over and above the effects of word recognition and language measures. The authors recommended modifying the simple view of reading (i.e., language and word recognition skill) to include fluency as a third component.

The importance of prerequisite reading skills to comprehension, as evidenced by the research cited previously, supports our decision to include prerequisite skill in the CBAL reading competency model. The competency model currently includes three types of prerequisite reading skills: the ability to recognize familiar words, the ability to decode unfamiliar words, and the ability to read connected text fluently. Because reading connected text is a more integrated task, such tasks are the preferred format for CBAL summative assessments. Similarly, because measures of decoding and word recognition skills can facilitate analyses of readers' underlying strengths and weaknesses, less integrated formats, such as asking students to read aloud from a list of pseudowords, are the preferred format for a more diagnostic or formative assessment. As a final note, it is important to reiterate that the construct of prerequisite reading skill refers to only the process of recognizing or decoding printed text; it does not refer to the general processes involved in comprehension.[5]

Recognize Technology's Impact: Reading in New Places

Currently, many standardized and accountability assessments of reading use traditional forms of printed material to assess reading comprehension (e.g., excerpts from novels, textbooks, newspapers, magazines). However, in the new age of informationalism (Castells, 1996), students are frequently called upon to participate in reading activities that involve technology (Kirsch et al., 2007; Partnership for 21st Century Skills, 2004, 2008; Warschauer, 2007). These activities include (but are not limited to) producing electronic documents, presentations, and communications (e.g., e-mail); performing information searches (e.g., Web browsing); and organizing electronic information. The use of technology is not limited to the work force, it extends into leisure time as well. More and more adolescents (and adults) are text messaging, reading blogs, engaging in online gaming and chat room discussions, watching videos (e.g.,

YouTube) and making connections through on-line social networking environments (e.g., Facebook). For example, according to the Pew Internet and American Life Project, a large proportion of people between the ages of 12 and 49 frequently use the Internet on a regular basis and this trend is increasing over time (for a brief overview of the project findings see, Jones & Fox, 2009). In order to accommodate these changes, CBAL assessments will utilize both traditional forms of printed material as well as the new forms of electronic media described previously.

Summary and Conclusions

This paper lays the foundation for a new approach for assessing reading comprehension in accountability settings. This new approach is called cognitively based assessment *of*, *for* and *as* learning (CBAL) because a cognitive model of domain competency provides the basis both for measuring learning and advancing learning. The approach involves first developing a cognitive model of reading competency and then using that model to design tasks and activities that provide high quality evidence about students' mastery status on targeted skills.

We began the discussion by emphasizing the need to test higher level reading skills such as integration/synthesis, evaluation/critical thinking, and explanation/reasoning. Collectively we refer to these skills as applied comprehension skill because they require readers to go beyond the basic meaning of a text in order to apply what has been learned in some meaningful way (e.g., solve a problem, make a decision). We outlined three reading activities in which applied comprehension skills are critically needed. The first activity involves constructing a document model to integrate and synthesize information from multiple texts. The second activity involves developing a critical stance in order to evaluate the quality of information. The third activity involves generating explanations both to develop deeper, more complete understandings of text content and to identify breakdowns in comprehension so that needed repairs can be made. A key aspect to the design involves presenting the assessment within the context of an authentic scenario designed to provide a realistic purpose for reading. Arguably, the construct of applied comprehension may go beyond general definitions of comprehension to include critical thinking, critical reasoning, decision making, and problem solving.

While the applied comprehension skills discussed previously are critically needed to succeed at 21st century tasks, they are difficult to master and breakdowns can occur at many levels. Consequently, we also advocate measuring two additional sets of skills: model building

skill and prerequisite reading skill. Model building skill is the collection of abilities needed to extract the basic meaning, or gist, of a text. This set can be further broken down into four components: understanding the literal meaning of a text (e.g., locate/retrieve), understanding key vocabulary, drawing inferences/ making generalizations, and extracting the discourse structure (organize/ summarize). We proposed graphical task formats designed to measure all three of these model building skills in an integrated manner because they closely align with strategies designed to improve reading comprehension. We also mentioned that model building skill could occur in the context of printed text or nonprinted formats (e.g., spoken language).

Because the ability to form an accurate mental model of a text depends on lower level reading skills, we also proposed several measures of prerequisite reading skill. In skilled readers, prerequisite reading skill requires little or no attention and resources. As a result, skilled readers are able to focus their attention and direct resources in ways that facilitate the application of higher level skills. On the other hand when lower level reading skills are not automated, valuable resources are no longer available for higher level processing. Consequently, we also advocated measuring three types of prerequisite reading skills: word recognition, decoding, and the ability to read text fluently. Prerequisite reading skill is bound to printed text and represents reading ability in its literal sense (unlike model building and applied comprehension, which can extend to forms of communication other than print).

Taken together, applied model building and prerequisite reading skill define a complex reading construct. Assessments that measure all three levels of this construct provide a powerful tool for assessing reading development and helping students to succeed in the 21st century.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge, New York: Cambridge University Press.

Anderson, L., & Krathwohl, D. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Allyn-Bacon Longman.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching* (pp. 77–117). Newark, DE: International Reading Association.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia. *Journal of Educational Psychology, 96*, 523–535.

Beck, I. L., & McKeown, M. G. (1991). Conditions of vocabulary acquisition. In R. Barr, M. L. Karmil, P. B. Mosenthal, & D. D. Pearson (Eds.), *Handbook of reading research: Vol. II* (pp. 749–814). New York: Longman Publishing Group.

Beck, I., McKeown, M., & Gromoll, E. (1989). Learning from social studies texts. *Cognition and Instruction, 6*, 99–158.

Bennett, R. E., & Gitomer, D. H. (in press). Transforming K–12 assessment. In C. Wyatt-Smith & J. Cumming (Eds.), *Educational assessment in the 21st century*. New York: Springer.

Bloom, B. S. (1956). Taxonomy of educational objectives, handbook I: The cognitive domain. New York: David McKay Co. Inc.

Bråten, I., & Samuelstuen, M. S. (2004). Does the influence of reading purpose on reports of strategic text processing depend on students' topic knowledge? *Journal of Educational Psychology, 96*, 324–336.

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes, 32*, 191–213.

Britt, M. A., & Aglinskas, C. (2002). Improving student's ability to use source information. *Cognition and Instruction, 20*(40), 485–522.

Britt, M. A., Goldman, S. R., & Perfetti, C. A. (1999, August). *Content integration in learning from multiple texts*. Paper presented at the meeting of the Society for Text and Discourse, Vancouver, BC, Canada.

Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology, 25*(4), 313–339.

Brown, A., & Campione, J. C. (1996). Psychological theory and the design of innovative learning assessments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Lawrence Erlbaum Associates.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and the relation to knowledge. *Memory & Cognition, 29,* 850–859.

Carver, R. P., & David, A. H. (2001). Investigating reading achievement using a causal model. *Scientific Studies of Reading, 5,* 107–140.

Castells, M. (1996). *The rise of network society*. Malden, MA: Blackwell.

Cataldo, M., & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology, 92,* 791–799.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Chi, M. T. H., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Cox, K. E., & Guthrie, J. T. (2001). Motivational and cognitive contributions to students' amount of reading. *Contemporary Educational Psychology, 26,* 116–131.

Cunningham, A., Stanovitch, K., & Wilson, M. (1990). Cognitive variations of adult college students differing in reading ability. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skill approaches* (pp. 129—159). New York Academic.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934–945.

Cutting, L., & Scarborough, H. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277–299.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (No. NCES 2006-469). Washington, DC: Government Printing Office.

Daneman, M. (1988). Word knowledge and reading skill. In M. Daneman, G. E. Mackinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 6, pp. 145–175). San Diego, CA: Academic Press.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422–433.

Deane, P., & O'Reilly, T. (2009). *Cognitively-based approach for measuring language arts skills*. Manuscript in preparation.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515–540.

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Trauber, E. R. (2003, June). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. In J. Arnowitz, A. Chalmers, & T. Swack (Chairs), *Proceedings of the conference on designing for user experiences* (pp. 1-15), New York: ACM.

Franzke, M., Kintsch, E., & Caccamise, D. (2005). Summary Street: Computer support for comprehension and writing. *Journal of Educational Computing Research, 33,* 53–80

Fritch, J. W., & Cromwell, R. L. (2002). Delving deeper into evaluation: Exploring cognitive authority on the Internet. *Reference Services Review, 30*(3), 242–254.

Garner, R. (1985). Text summarization deficiencies among older students: Awareness or production ability? *American Educational Research Journal, 22,* 549–560.

Garnham, A., Oakhill, J. V., & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. *Cognition, 11,* 29–46.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Glenberg, A., Wilkinson, A., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10,* 597–602.

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Ferris & D. M. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 317–351). Greenwich, CT: Information Age Publishing.

Goldman, S. R., & Bisanz, G. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 19–50). Mahwah, NJ: Lawrence Erlbaum Associates.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Gough, P. B., & Walsh, M. A. (1991). Chinese, Phoenicians, and the orthographic cipher of English. In S. A. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 199–209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist, 40,* 225–234.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31,* 104–137.

Graesser, A. C., Wiley, J., Goldman, S., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK Web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2,* 89–105.

Graves, M. F., Brunetti, G. J., & Slater, W. H. (1982). The reading vocabularies of primary-grade children of varying geographic and social backgrounds. In J. A. Niles & L. A. Harris (Eds.), *New inquiries in reading research and instruction: Thirty-first yearbook of the National Reading Conference* (pp. 99–104). Rochester, NY: National Reading Conference.

Greene, S. (1994). Students as authors in the study of history. In G. Leinhardt, I. Beck, & K. Stainton (Eds.), *Teaching and learning history* (pp. 133–168). Hillsdale, NJ: Lawrence Erlbaum Associates.

Griffin, C., & Tulbert, B. (1995). The effect of graphic organizers on students' comprehension and recall of expository text: A review of the research and implications for practice. *Reading and Writing Quarterly*: *Overcoming Learning Difficulties, 11*, 73–89.

Guthrie, J. T., & Mosenthal, P. (1987). Literacy as multidimensional: Locating information and reading comprehension. *Educational Psychologist, 22*, 279–297.

Haenggi, D., & Perfetti, C. (1994). Processing components of college-level reading comprehension. *Discourse Processes, 17*, 83–104.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing Co.

Head, M., Readence, J., & Buss, R. (1989). An examination of summary writing as a measure of reading comprehension. *Reading-Research-and-Instruction, 28,* 1–11

Hirsch, E. D. (2003, Spring). Reading comprehension requires knowledge of words and the world. *American Educator*, pp. 10–31.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160.

Jenkins, J., Fuchs, L., van den Broek, P., Espin, C., & Deno, S. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719–729.

Jones, S., & Fox, S. (2009). *Generations online in 2009* [PEW Internet Project Data Memo]. Retrieved March 20, 2009, from http://www.pewinternet.org/~/media//Files/Reports/2009/PIP_Generations_2009.pdf.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future* (Policy Information Rep.). Princeton, NJ: ETS.

Lambiotte, J., Skaggs, L., & Dansereau, D. (1993). Learning from lectures: Effects of knowledge maps and cooperative review strategies. *Applied Cognitive Psychology, 7*, 483–497.

Langer, E. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Linderholm, T., Virtue, S., Tzeng, Y., & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes, 37,* 165–186.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1456–1470.

Lorch, R. F., Jr., Lorch, E. P., & Morgan, A. M. (1987). Task effects and individual differences in on-line processing of the topic structure of a text. *Discourse Processes, 10*, 63–80.

Lorch, R., Pugzles-Lorch, E., Ritchey, K., McGovern, L., & Coleman, D. (2001). *Contemporary-Educational-Psychology, 26*, 171–191.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21*, 251–283.

Magliano, J. P., Todaro, S. Millis, K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research, 32*, 185-208.

Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Muñoz, B. D., & McNamara, D. (2002). Using latent semantic analysis to assess reader strategies. *Behavior Research Methods, Instruments & Computers, 34*, 181–188.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 11–151.

McCormick, S. (1994). A nonreader becomes a reader: A case study of literacy acquisition by a severely disabled reader. *Reading Research Quarterly, 29*, 156–177.

McCutchen, D., & Perfetti, C. (1982). Coherence and connectedness in the development of discourse production. *Text, 2*, 113–139.

McGlinchey, M. T., & Hixon, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 193–203.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440–466.

McNamara, D. S. (2004). SERT: Self explanation reading training. *Discourse Processes, 38*, 1–30.

McNamara, D. S., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1–43.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers, 36*, 222–233.

McNamara, D. S., & O'Reilly, T. (in press). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In F. Columbus (Ed.), *Progress in experimental psychology research*. Hauppauge, NY: Nova Science Publishers, Inc.

McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34*, 147–171.

McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). A reading strategies framework. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 465–496). Mahwah, NJ: Lawrence Erlbaum Associates.

McNeil, L. M. (2000). Contradictions of school reform: Educational costs of standardization testing. New York: Routledge.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*(13), 2078–2091.

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D., & McCann, R. (2003). Bringing the concept of credibility into the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication Yearbook, 27*, 293–335.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16*, 72–103.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of the structure strategy and signaling on recall of the text. *Journal of Educational Psychology, 93,* 141–159.

Meyer, B., & Wijekumar, K. (2007). A Web based tutoring system for the structure strategy: theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 347–374). Mahwah, NJ: Lawrence Erlbaum Associates.

Millis, K., Kim, H.J., Todaro, S., Magliano, J.P., Wiemer-Hastings, K., & McNamara, D.S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, & Computers, 36,* 213-221.

Moore, S. R. (1995). Focus on research questions for research into reading-writing relationships and text structure knowledge. *Language Arts, 72,* 598–606.

Nagy, W., & Scott, J. (2000). Vocabulary process. In M. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 269–284). Mahwah, NJ: Lawrence Erlbaum Associates.

Nicholas, S. L., & Berliner, D. C. (2007). *Collateral damage: How high stakes testing corrupts America's schools*. Cambridge, MA: Harvard Educational Press.

No Child Left Behind Act of 2001, 20 U.S.C. § 6301 *et seq.* (2002).

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology, 54,* 31–39.

Oakhill, J., & Yuill, N. (1986). Pronoun resolution in skilled and less-skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech, 29,* 25–37.

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69–92). Mahwah, NJ: Lawrence Erlbaum Associates.

Oakhill, J. V., Yuill, N., & Donaldson, M. L. (1990). Understanding of causal expressions in skilled and less skilled text comprehenders. *British Journal of Developmental Psychology, 8,* 401–410.

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum Associates.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3*(4), 229–235.

Ouellet, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566.

Ozuru, Y., Best, R., & McNamara, D. S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th*

annual meeting of the Cognitive Science Society (pp. 1071–1076). Mahwah, NJ: Lawrence Erlbaum Associates.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40,* 1001–1015.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117–175.

Partnership for 21st Century Skills. (2004). *Learning for the 21st century: A report and mile guide for 21st century skills.* Retrieved May 11, 2009, from http://www.21stcenturyskills.org/images/stories/otherdocs/p21up_Report.pdf

Partnership for 21st Century Skills. (2008). *21st century skills and English map.* Retrieved May 11, 2009, from http://www.21stcenturyskills.org/documents/21st_century_skills_english_map.pdf

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education..

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Philadelphia: John Benjamins.

Perfetti, C. A., Rouet, J., & Britt, M. A. (1999). Toward a theory of document representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Lawrence Erlbaum Associates.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III* (pp. 545–561). Mahwah, NJ: Lawrence Erlbaum Associates.

Radmacher, S., & Latosi-Sawin, E. (1995). Summary writing: A tool to improve student comprehension and writing in psychology. *Teaching of Psychology, 22*, 113–115.

Rand Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: Author.

Rinehart, S. D., Stahl., S. A., & Ericson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly, 21,* 422–437.

Robinson, D., & Skinner, C. (1996). Why graphic organizers facilitate search processes: Fewer words or computationally efficient indexing? *Contemporary Educational Psychology, 21*, 166–180.

Rowe, D. (1994). *Preschoolers as authors: literacy learning in the social world of the classroom.* Cresskil, NJ: Hampton Press.

Roy, M., & Chi, M. (2005). The self-explanation principle in multi-media learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 271-286). Cambridge, New York: Cambridge University Press.

Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23,* 441–474.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during internet research tasks. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the 7th international conference of the learning sciences* (pp. 662–666). Bloomington, IN: International Society of the Learning Sciences.

Scholz-Crane, A. (1998). Evaluating the future: A preliminary study of the process of how undergraduate students evaluate Web sources. *Reference Services Review*, *26,* 53–60.

Schraw, G., & Dennison, R. S. (1994). The effect of reader purpose on interest and recall. *Journal of Reading Behavior, 26,* 1–17.

Sheehan, K. M., Kostin, I., & Futagi, Y. (2008a, July). *Reading level assessment for high-stakes testing applications: A second look at variation due to differences in text genre*. Paper presented at the annual meeting of the Society for Text and Discourse, Memphis, TN.

Sheehan, K. M., Kostin, I., & Futagi, Y. (2008b, July). When do standard approaches for measuring vocabulary difficulty, syntactic complexity and referential cohesion yield biased estimates of text difficulty? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1978-1982). Austin, TX: Cognitive Science Society..

Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language*, *33,* 421–441.

Singer, M., & Remillard, G. (2004). Retrieving text inferences: Controlled and automatic influences. *Memory & Cognition, 32*, 1223–1237.

Smith, J., & Niemi, R. (2001). Learning history in school: The impact of course work and instructional practices on achievement. *Theory and Research in Social Education, 29*, 18–42.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56,* 72–110.

Stevens, R. J. (1988). Effects of strategy training on the identification of the main idea of expository passages. *Journal of Educational Psychology, 80*, 21–26.

Tannenbaum, R., Torgesen, J., & Wagner, R. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 381–398.

Thiede, K., & Anderson, M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28,* 129–160.

van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29,* 1081–1087.

van den Broek, P., Risden, K., & Husebye-Hartman, E. (1995). The role of the reader's standards of coherence in the generation of inference during reading. In R. F. Lorch, Jr., & E. J. O'Brian (Eds.), *Sources of coherence in text comprehension* (pp. 353–373). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vansledright, B. (1995). "I don't remember—The ideas are all jumbled in my head": Eighth graders' reconstructions of Colonial American History. *Journal of Curriculum and Supervision, 10*, 317–345.

Vansledright, B., & Kelly, C. (1998). Reading American history: The influence of multiple sources on six fifth graders. *The Elementary School Journal, 12*, 1–18.

Vellutino, F., Tunmer, W., & Jaccard, J. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3–32.

Vitale, M., & Romance, N. (2007). A knowledge-based framework for unifying content-area reading comprehension and reading comprehension strategies. In D. S. McNamara (Ed.),

Reading comprehension strategies: Theory, interventions, and technologies (pp. 73–104). Mahwah, NJ: Lawrence Erlbaum Associates.

Wang, J. H., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students. *Reading Research Quarterly, 39,* 162–186.

Warschauer, M. (2007). Information literacy in the laptop classroom. *Teachers College Record, 109*, 2511–2540.

Wiley, H., & Deno, S. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* 207–214.

Wiley, J., & Voss, J. (1996). The effects of playing historian on learning in history. *Applied Cognitive Psychology, 10*, 63–72.

Williams, A. (1986). Teaching children to identify the main idea of expository texts. *Exceptional Children, 53*, 163–168.

Williams, J. (2007). Literacy in the curriculum: Integrating text structure and content area instruction. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 199–219). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, P. T., & Anderson, R. C. (1986). What they don't know will hurt them: The role of prior knowledge in comprehension. In J. Orasanu (Ed.), *Reading comprehension: From research to practice* (pp. 31–48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science, 22*, 319–346.

Yuill, N., & Oakhill, J. (1988a). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology, 79*, 173–186.

Yuill, N., & Oakhill, J. (1988b). Effects on inference awareness training of poor reading comprehension. *Applied Cognitive Psychology, 2,* 33–45.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology, 80*, 351–361.

Notes

[1]For example, one design might specify using four 90-minute summative assessments to be administered at four different time points throughout the year for a total of 360 minutes of testing time. As a result, each 90-minute assessment could be more focused and therefore assess the standards and necessary skills at a greater depth than what would be possible in a typical one-time year-end assessment.

[2] Appendix A of this report describes the skills branch of the competency model in more detail.

[3] Collectively, the skills branch has some overlap with Bloom's Taxonomy that is often used in educational settings (Anderson & Krathwohl, 2001; Bloom, 1956). While differences exist in the models, Bloom's notions of knowledge and comprehension roughly correspond to model building skill, while Bloom's notions of application, analysis, synthesis, and evaluation roughly correspond to applied comprehension skill. A detailed comparison between Bloom's taxonomy and the CBAL reading and writing competency models is found in Deane and O'Reilly's work (2009).

[4] We prefer to use the term *discourse structure* as opposed to *text structure* to reflect the generative nature of comprehension. Text structure is often used to describe the features of the text. In contrast, discourse structure is intended to refer to the structure of the reader's representation of the text. This is a very subtle but important difference.

[5] Although we recognize that the skills underlying higher level comprehension can also affect lower level processing such as word recognition.

[6] The individual competencies included in this set are the observable manifestations of the latent construct of automaticity, characterized cognitively as requiring minimal cognitive effort and attention. The performance appears as if a direct memory trace between spelling and meaning and/or pronunciation has been forged, though this latter theoretical claim is difficult to know or prove.

Appendix

Cognitively Based Assessment *of*, *for* and *as* Learning (CBAL) Competency Model—The Skills Branch

This section describes the skills branch of the CBAL reading competency model. It is broken down into three areas: prerequisite reading skill, model building skill, and applied comprehension skill. The top portion of Figure 1 is a graphical depiction of the skills branch of the CBAL reading competency.

Prerequisite Reading Skill

Prerequisite reading skill is the ability to accurately read text at a rate that is grade level appropriate. It includes the ability to recognize familiar words (e.g., house) and the ability to decode unfamiliar real or pseudowords. Prerequisite reading skill includes the following components:

Recognize Common or Familiar Words

This skill involves rapidly recognizing grade level appropriate familiar words. For students without disabilities, the skill involves recognizing words in printed or digital text. The recognize common or familiar words skill category differs from the other abilities listed in that only minimal processing effort is involved.[6]

Decode Unfamiliar Words or Pseudowords From Text

This skill is the ability to produce plausible pronunciations of unfamiliar or pseudowords by applying English orthographic/phonetic conventions.

Read Connected Text Fluently and With Appropriate Prosody

This is the ability to read aloud at a rate and level of accuracy that is grade level appropriate. In some cases the construct may include the concept of prosody which is defined as the appropriate use of intonation, rhythm, focus, and emphasis where appropriate.

Model Building Skill

Model building skill concerns the reader's ability to form a gist representation of the information contained within a single passage. The construct is broken down into four components: understanding the literal meaning (e.g., identifying important details,

locate/retrieve), understanding grade appropriate vocabulary, drawing necessary inferences/making key generalizations, and extracting the discourse structure (chunk, organize, summarize).

Understanding the Literal Meaning of a Passage (Locate/Retrieve)

This category concerns the ability to locate/retrieve, remember, and understand the details, facts, concepts, and individual sentences contained within a passage. This skill category also includes comprehending/remembering the gist of important sentences, and recognizing and producing accurate paraphrases of a sentence. Questions designed to measure mastery status relative to these skills will not require inferences or background knowledge. All of the necessary information required to answer literal questions is contained within the text. For the purposes of item writing, this construct is broken down into two components: the ability to locate and retrieve information, and the ability to produce and recognize accurate paraphrases of facts, concepts, and individual sentences contained within a passage.

Understand Grade Appropriate Vocabulary

This component includes the ability to acquire and use the type of technical vocabulary typically found in grade level appropriate text. It includes the ability to infer the meaning of unknown words from the contexts in which they are used and the ability to use acquired vocabulary in meaningful and appropriate ways.

Infer and Generalize Information (Infer Relations Among Concepts or Text-Based Inferences)

This component concerns the ability to infer the relations among the concepts expressed within a single text. It includes the ability to make inferences across adjacent sentences, as well as the ability to make inferences across distal sentences. The construct also includes the ability to extract themes and main ideas from text. All the information required to answer this item type is provided within the text. No special background knowledge is necessary. As a general rule, items that require the use of two or more sentences are considered inference items, and items that require the use of background knowledge are explanation based or explanatory inference items (i.e., an applied comprehension construct). There are two basic types of inference items: local and global inference. The simplest form of a local inference is the anaphoric type. Anaphoric inferences involve making connections among local ideas in text by using discourse markers

(e.g., he, she, they, it, latter, second) to connect ideas across sentences. Another, more demanding type of local inference requires the reader to make inferences that extend the explicit situation depicted in the text. The second basic type of inference is the global or thematic inference. This type of inference requires the reader to make generalizations based on large portions of text.

Extract and Use Discourse Structure (Chunk, Organize, Summarize)

This component concerns the ability to formulate a network representation of how the main ideas and supporting details are related to each other. Discourse structure involves identifying the complex spatial, temporal, and causal relationships that summarize and classify the key relationships (often hierarchical) among the core ideas in text. For literary texts, discourse structure involves identifying the relationship between characters, plot, setting, conflicts, and resolutions, and distinguishing chronological sequence from the narrative sequence. For expository texts, discourse structure involves classifying, listing, describing, sequencing, comparing, and contrasting concepts; identifying the problem and solutions; and identifying cause and effect relationships. By definition, discourse structure is an amalgamation of both literal and inferential information that results in a succinct, orderly synopsis of the concepts and the relationships among ideas in text. For some texts the discourse structure is relatively explicit and straightforward. For other texts the structure is implicit and has to be inferred.

Applied Comprehension Skill

Evaluate and Critique (Critical Reading/Thinking)

This component concerns the ability to deploy critical thinking skills to evaluate text. Evaluation involves judging, ranking, and appraising the value of a text as a whole and the specific features that comprise it. It also involves verifying the truth value of text contents where appropriate (e.g., if the student has adequate knowledge of the topic). In narrative text, it involves the ability to evaluate author craft. In persuasive text, it involves the ability to critically evaluate author/ source creditability, claims, support for claims and conclusions, implications of claims and conclusions, and how current the information is. It includes the ability to detect the author's purpose for writing, position on the issue, potential bias, logical flaws in the arguments, and the ability to identify conflicting arguments. It also includes the ability to determine how

relevant an argument or document is for the issue at hand and the how relevant it is for achieving the reader's overall goal. In short, evaluation involves discerning the quality of text content as defined by the six dimensions mentioned in the prior review: accuracy, authority, objectivity, currency, coverage, and relevance.

Integrate and Synthesize With Other Text

This component concerns the ability to understand, integrate, and synthesize information from multiple documents. It is the ability to manage multiple documents by combining information from each document to produce a coherent, meaningful, and systematic document model of the text content (i.e., integrate). In other words, integrating requires the reader to identify and combine information from several documents that is relevant for achieving his or her overall purpose for reading. It also involves the ability to synthesize, which is defined as the ability to generate new ideas, concepts by combining the information from two or more texts in novel ways. While constructing a document model requires the use of similar skills involved in building a model of a single text (model building), the process involved in creating a document model is much more complex. Within a single text, there is often an explicit or strongly implied general model of the text; that is, the ideas contained within a single text should be logically connected and the text as a whole should be coherent. This is because the author wrote the text for a particular purpose and audience. In contrast, a document model requires the reader to notice connections, draw comparisons, and make contrasts between two or more texts that are often written by different authors and for different purposes. Such differences make the process of document building much more difficult because the links among the documents are not explicitly provided, and the reader has to generate and infer the links based on the reader's purpose for reading the collection of documents. The construct also involves not only linking documents, but also the ability to apply the principles and concepts learned in one or more documents to hypothesize, extrapolate, and solve problems contained in related texts. This application of core ideas or transfer of knowledge is one of the hallmarks of expert performance.

Generate Explanations—The Ability to Reason and Explain Text

This component concerns the ability to construct a deep and consistent understanding of text by generating explanations. The construct includes the ability to draw inferences to fill in missing information, to make connections among ideas that are not easily inferable, to integrate

text information with existing knowledge, and to monitor and repair incorrect knowledge. Generating explanations involves the ability to use general background knowledge to elaborate, explain, reason, or infer unstated relationships within text. Explaining involves the ability to monitor comprehension for the purposes of identifying inconsistencies in text and also to identify misconceptions in understanding. Generating explanations often involves asking and answering deep questions that focus on causes, motivations, supporting evidence, planning, and logical justifications. Such questions can take on many generic forms, such as the following: Why? What caused X? How did X occur? What if, what if not? How does X compare to Y? What is the evidence for X? Why is X important? Generating explanatory inferences is different from the draw inferences competency included under the model building skill (i.e., text-based inferences). Generating explanations requires the use of general knowledge and reasoning processes to connect concepts, while text-based inferences can be made on the basis of the information contained solely within the text.